PROCEDURAL INNOVATIONS IN REVISIONARY STUDIES: COMPUTER-ASSISTED CITATION OF REPRESENTATIVE SPECIMENS

Christopher A. Meacham
Department of Botany, University of Michigan, Ann Arbor 48104
Tod F. Stuessy
Department of Botany, Ohio State University, Columbus 43210

Revisionary studies are in many ways the cornerstones of systematic botany. It is within revisions that basic relationships among specific plant groups are sketched, and it is within this framework, therefore, that most data regarding classification and evolution of plants have been documented. The importance of revisionary studies in plant systematics has been emphasized with some regularity (e.g., Robinson, 1923; Just, 1954; Stuessy, 1975).

Although presently used, time-tested methods for producing revisions have been highly successful, one wonders if, with procedural modifications, revisions could be produced more efficiently and with greater accuracy of cited information. If this were so, then the main contributions of revisionary studies to plant systematics would become even more significant, and we might come more quickly to an understanding of the diversity and interrelationships of the whole plant kingdom. This is particularly significant in view of the rapid disappearance of much of the world's flora due to the increasing size and activities of the human population.

A number of possibilities exist for procedural innovations in revisionary studies, but perhaps one of the most significant changes will come with the use of data-processing machines. The utility of a revision relates to the ease with which information can be retrieved from the revisionary framework of keys, descriptions, representative specimens, and distribution maps. The procedures involved in generating these data and presenting them in standardized formats are conceptually simple, laborious, and error-prone. It is precisely this set of attributes that is well-suited for computer application. With explicit instructions, computers can perform simple, routine tasks very rapidly and with few, if any, errors.

Although several projects already have utilized computers in procedures that relate to revisionary studies, e.g., in automatic key generation (Hall, 1970; Morse, 1971, 1974; Pankhurst, 1971, 1974; Pettigrew and Watson, 1973; Dallwitz, 1974), in automatic preparation of distribution maps (Soper, 1964, 1969; Adams, 1974), and in extracting data from floras (Keller and Crovello, 1973), none has employed the computer to produce lists of representative specimens in conventional format. Argus and Sheard (1972), Argus (1973), and Shetler et al. (1973) have used the computer to generate lists of specimens, but these are not concatenated in the standard format.

Representative specimen citations serve three important functions in revisionary studies: (1) they indicate at least some of the specimens seen or studied by the author; (2) they include information on precise locality, date of collection, and name of collector(s), which is of interest to many people, such as ecologists, floristic workers, or biographers; and (3) they serve as a convenient catalogue of identified specimens so that floristic workers or herbarium curators can refer to these data and identify duplicate material without having to write to the specialist for aid.

In our laboratory we have used computers to prepare specimen citation lists for revisionary studies on the subtribe Lagasceinae (Compositae, Heliantheae). These procedures have been helpful to our work, and we present this paper with the following specific purposes: (1) to describe a program for computer-assisted citation of representative specimens (called REVISO) and (2) to comment on the time, cost, and accuracy of using this new method with more conventional manual-file procedures. We realize that several large electronic data-processing systems such as SELGEM at the U.S. National Museum, the Generalized Information System of the now dormant Flora North America Program (Krauss 1973), the program package used by Crovello (1972) for the Greene Herbarium, or TAXIR, developed at the University of Colorado (Brill, 1971), could be modified to accomplish what we have done with REVISO (see Crovello and MacDonald, 1970, and Brenan, 1974, for a more complete index of electronic data-processing programs available). Even if this were done, however, we believe that the availability of a relatively simple, short, inexpensive program is more attractive to practicing taxonomists in their own institutions than is the prospect of time-sharing with one of the big systems (see Argus and Sheard, 1972, for a similar viewpoint).

Description of Program (REVISIO)

The REVISIO program is designed to input, manipulate, and output data consisting of label information from herbarium specimens. It is written in the PL/I optimizer language and requires 252K of storage. The types of manipulations and formats of output are determined by user-supplied control cards. The following description of the functions performed by REVISIO is intended only as a general overview.

The operation of REVISIO can be divided conceptually into three phases: data input, data manipulation, and data output. Each of these operations is discussed below.

Data Input

In the data input phase, information is read from the input file, edited to some extent, and placed in an array in storage. For most users, the input file will be a deck of cards. It is useful to examine the format of the card containing the label information in some detail, because this

will provide a good basis for understanding the operation of the program.

Each set of information from one label (hereafter called an "entry") is punched onto one or more cards. Each entry consists of six fields separated by asterisks: geography*locality*date*collector*number* herbaria*. Six asterisks must appear in each entry. The taxon designation for each of the entries is not put on the entry card itself but is provided by an additional card that precedes all entries for that taxon.

Geography.—The geography field consists of names of the political subdivisions where the specimen was collected. These names may be written in full or may be abbreviated. Full names are followed by semicolons; abbreviations, by colons. For example: US:OH:FRANKLIN CO.;*. A list of abbreviations and their full equivalents supplied to the program by the user enables the program to replace these abbreviations on output. Abbreviations do not have to be used, however, but their use can save time in data input if many specimens are from the same region.

Locality.—The specific locality is simply keypunched as it occurs on the label (often with minor editorial modifications), including brief ecological or elevational data if desired.

Date.—The date can be keypunched in several ways: 1904*; MAY 1904*; 8MAY1904*; 08 MAY 1904*; or "WITHOUT DATE"*.

Collector.—The name of the collector(s) is punched thus: MURPHY*; MURPHY ET AL.*; MURPHY & JONES*; or COLLECTOR UNKNOWN*.

Number.—The collector's number is entered in whatever form it appears on the label, consisting of numbers and/or letters, and including the use of *s.n.* when the number is not known.

Herbarium.—The last field is a list of the herbarium abbreviations (from *Index Herbariorum,* Holmgren and Keuken, 1974) separated by commas: GH, NY, F, US, OS*. The abbreviations may be input in any sequence; before data output they are sorted alphabetically.

An entire entry might look like this: US:ARI:PIMA CO.;*BABOQUIVARI MTS*16APR1935*MAGUIRE*10746*GH, NY*.

In some cases the information from the label will not fit into the various formats just described. For example, in the case of dates a range may be specified: 16–20 JAN 1895. To deal with this type of problem, REVISIO has the capability of storing one piece of information for manipulation while using the other for output. On the data card the manipulation datum is keypunched first in the date field followed by what is to be printed in quotes: 16JAN1895 "16–20 JAN 1895"*. If a sort by date is requested, this entry will be sorted with a date of 16 JAN 1895, but a listing will produce a date of 16–20 JAN 1895. This substitute option is available for three of the fields: geography, date, and number.

Taxon Name and Number.—In addition to the information on the specimen label, each entry is associated with a taxon name and its arbitrarily assigned number. This taxon information is read from the

card that precedes the entries for that taxon. The taxon name and number may be sorted during data manipulation as though they were keypunched individually for each entry.

Thus, after input, there are eight kinds of information stored for each entry, six from the label proper, plus one for the taxon name and another for an arbitrarily assigned taxon number.

Data Manipulation

During data manipulation, there are two functions available: the INCLUDE-EXCLUDE (IN-EX) function, and the SORT function. The IN-EX function allows the user to specify a portion of the entire data set for output. The user may specify, for example, a state name, a collector, or a range of dates. By using a number of INCLUDE's or EXCLUDE's even more specific groups may be defined. An INCLUDE Ohio, EXCLUDE Franklin Co., INCLUDE Murphy, INCLUDE dates >1950, will produce a listing of all the entries for collections made by Murphy in Ohio, outside of Franklin Co., after 1950.

The SORT function enables the user to sort the data set or selected subset prior to output. The user supplies a control card indicating which fields should be sorted and in what order. A common choice would be first by taxon; then within taxon by collector; and finally within collector by number. The control card for this sort would be: TX CL CN.

Data Output

After manipulating the data, the TABLE and REPRESENTATIVE SPECIMEN functions cause the data set or selected subset to be listed. When using the TABLE function, the user furnishes a format card that directs REVISIO to list the requested information in columns. As an example, the format card, CL(10) DT(12) G1(15) G2(15) LC(30), will allot the first ten print positions to collector, the second 12 to the date, the next 15 to the major geographic division (probably country), the next 15 to the first subdivision (state), and the next 30 print positions to specific locality. Any of the eight pieces of information included in the entry may be requested and listed in this manner.

The REPRESENTATIVE SPECIMEN function causes the data set or subset to be listed in paragraph form with correct punctuation as used for lists of representative specimens in revisionary studies.

The END function brings REVISIO to normal termination.

Description of Procedures with REVISIO

Although different types of procedures can be accomplished either separately or in sequence with REVISIO, we believe one particular series of procedures to be most helpful for use in revisionary investigations (fig. 1). The following paragraphs describe briefly some of these steps.

We have found it very useful to generate checklists of data in varying formats as an aid to the revisionary studies in progress. REVISIO not only will generate data in representative specimen format but also in

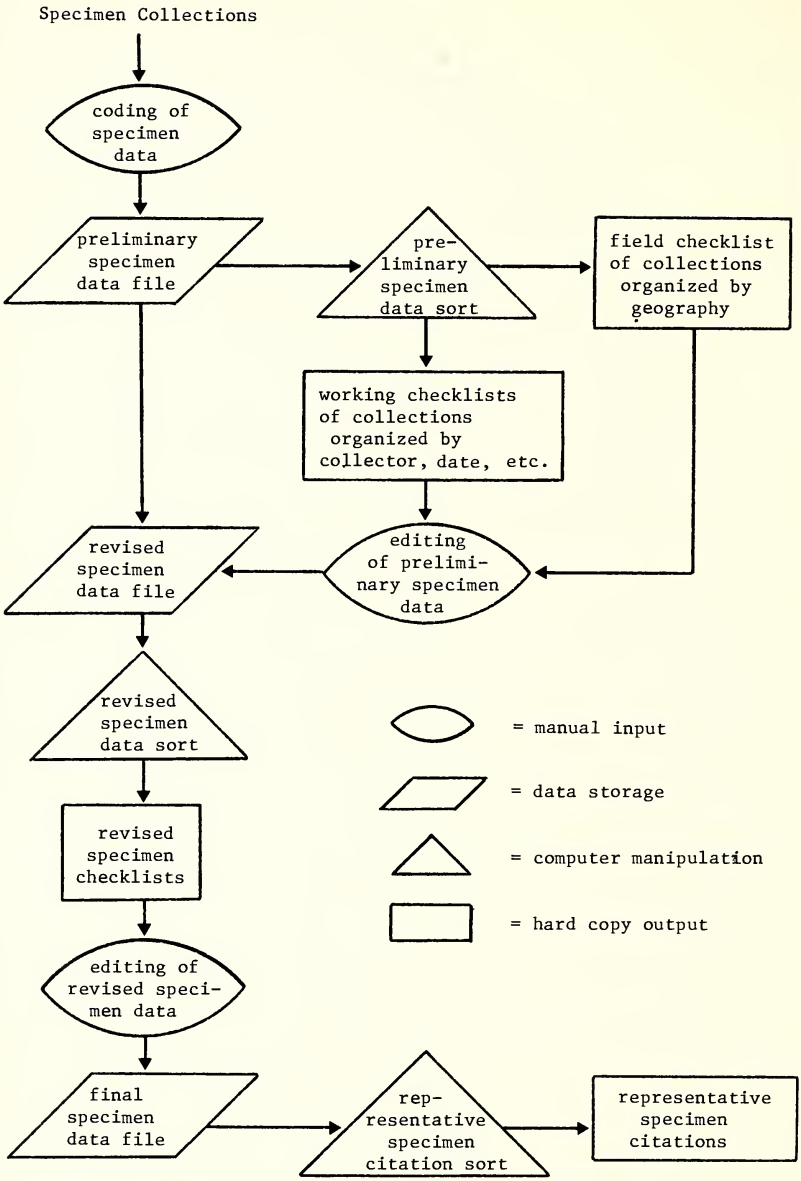

FIG. 1. Flow chart of procedures with REVISIO.

numerous other modes. For example, we have found it helpful to obtain a print-out of the specimen data that centers on the collector and his numbers, followed by the other information, for the following reasons: (1) locating type specimens in collections at hand; (2) detecting errors

in data, such as variant or erroneous spellings in collector's names; (3) locating duplicated collection numbers of the same collector; and (4) making editorial changes in the way data are cited, e.g., eliminating unnecessary locality data. We have also found a list of collections ordered geographically by region along a proposed itinerary very helpful in field studies.

After all collections have been obtained and data from these have been entered into the specimen data file, final representative specimen listings can then be generated. The representative specimen format in REVISIO is one that is commonly used by revisionary workers. The order of data in each citation is: geography, locality, date, collector, number, and herbaria. The sequence of citation is: (1) alphabetical by country and then by state and county; (2) alphabetical by the name of the collector(s); and (3) numerical in ascending order by the collection number. All the proper punctuation of periods, colons, semicolons, and commas are added automatically, and herbarium abbreviations are arranged alphabetically and enclosed with parentheses. The finished product (fig. 2) is indented in proper paragraph form and is printed out to fit on a page size 8½ inches wide. Use of cards for data submission results in all upper-case letters (as shown), whereas data input through a time-share terminal gives both upper and lower-case capability.

Use of REVISIO vs. Conventional Procedures

Interest on the part of practicing taxonomists to use REVISIO might depend upon knowing the costs of using the program in comparison with conventional manual card-file techniques. Three aspects need to be compared: time, money, and accuracy.

REPRESENTATIVE SPECIMENS. EL SALVADOR. AHUACHAPAN: NEAR ATACO, 19 JAN 1947, STANDLEY & PADILLA V. 2633 (F). SAN SALVADOR: SAN SALVADOR, JAN 1923, CALDERON 1456 (GH, NY). SAN VICENTE: NEAR SAN VICENTE, 2-11 MAR 1922, STANDLEY 21653 (GH). GUATEMALA. GUATEMALA: NEAR FINCA LA AURORA, 1938-39, AGUILAR 1 (F); VILLA CARLOTTA, 29 NOV 1896, SELER, C. & E. 2484 (GH, NY). JUTIAPA: JUTIAPA, WITHOUT DATE, MORALES R. 1362 (F). ZACAPA: SIERRA DE LAS MINAS, TRAIL ABOVE RIO HONDO, 11 OCT 1939, STEYERMARK 29549 (F). MEXICO. GUERRERO: 7.2 MI S OF CHILPANCINGO ON RTE 95, 11 SEP 1973, STUESSY & GARDNER 3146 (OS); 7.1 MI S OF PALO BLANCO ON RTE 95, 11 SEP 1973, STUESSY & GARDNER 3152 (OS).

Fig. 2. Example of a representative specimen listing produced by REVISIO.

It is somewhat difficult to compare directly the time involved with REVISIO versus traditional methods. Inherent in the difficulty is the usual lack of familiarity of most taxonomists or herbarium secretaries with a keypunch machine, whereas these same people are usually skilled with a typewriter. To put the comparison on equal grounds, if the user is skilled to a high degree on both machines, in our experience, use of the keypunch for data input does take a little more time than with a typewriter. However, the subsequent direct computer listing of representative specimens is much faster than having to type them out by hand. In general, we can say that if small amounts of data are involved, say, fewer than 100 collections, and no additional lists or sorts are wanted—only the representative specimen listing—then it probably is faster to operate with conventional techniques. However, if more than 100 collections are involved, and/or if many data manipulations are desired, then use of REVISIO should result in a net saving of time. Obviously, the more data involved and the greater the number of manipulations desired, the more time will be saved by using the computer.

If one does not include time in cost analysis, then the use of REVISIO is more expensive than conventional methods. Computer cards are available gratis or at minimal cost ($2/thousand), and keypunch machines are accessible at no charge at many institutions. Although use of the keypunch and cards is minimal, use of computer time is expensive. A computer run for a representative specimen printout with REVISIO will vary in cost from about five to ten dollars (15–20 seconds CPU), depending upon the size of the data set being used.

If one is careful in the initial input of data with REVISIO, the accuracy of the sorted and printed information is astonishing. Once the data have been punched error-free on the cards, the computer will perform countless manipulations with no mistakes. This is a decided advantage in dealing with representative specimen data, because errors can so easily occur in each retyping step with conventional procedures. Another advantage in using REVISIO is the opportunity to edit data several times in differing formats before the final representative specimen list is produced. This procedure helps eliminate errors that were inadvertently made at time of original data input.

In summary, if small amounts of data (100 collections or less) are being processed, or if money for computer time is not available, then use of REVISIO is not recommended. However, because most revisionary studies involve several thousands of collections, many taxonomists may find the program helpful. Computer operation costs must be met, of course, but many institutions have mechanisms for allowing time to be obtained, if a genuine need is shown. Perhaps the best way for taxonomists to learn if REVISIO is suited for their own studies, is to try it at their home institutions. Copies of the program, sample data, and a user's manual are available from the second author.

Acknowledgments

The Instructional and Research Computer Center of Ohio State University provided computer time for the development of REVISIO. Certain aspects of this study were supported by National Science Foundation Grant GB-37678. Thanks go to Drs. T. J. Crovello and E. D. Rudolph for critical evaluation of the final manuscript.

Literature Cited

Adams, R. P. 1974. Computer graphic plotting and mapping of data in systematics. Taxon 23:53–70.

Argus, G. W. 1973. The genus *Salix* in Alaska and the Yukon. Natl. Mus. Canada Publ. Bot., No. 2.

———. and J. W. Sheard. 1972. Two simple labeling and data retrieval systems for herbaria. Canad. J. Bot. 50:2197–2209.

Brenan, J. P. M. 1974. International conference on the use of electronic data processing in major European plant taxonomic collections. Taxon 23:101–107.

Brill, R. C. 1971. The taxir primer. Occas. Paper No. 1, Instit. Arctic & Alpine Research, Univ. Colorado, Boulder.

Crovello, T. J. 1972. Computerization of specimen data from the Edward Lee Greene Herbarium (ND-G) at Notre Dame. Brittonia 24:131–141.

——— and R. D. MacDonald. 1970. Index of edp-ir projects in systematics. Taxon 19:63–76.

Dallwitz, M. J. 1974. A flexible computer program for generating identification keys. Syst. Zool. 23:50–57.

Hall, A. V. 1970. A computer-based system for forming identification keys. Taxon 19:12–18.

Holmgren, P. K. and W. Keuken. 1974. Herbarium abbreviations. *In,* Index Herbariorum, part 1, ed. 6. Regnum Veg. 92:303–354.

Just, T. 1954. Generic synopses and their role in modern botanical research. Taxon 3:201–202.

Keller, C. and T. J. Crovello. 1973. Procedures and problems in the incorporation of data from floras into a computerized data bank. Proc. Indiana Acad. Sci. 82:116–122.

Krauss, H. M. 1973. The information system design for the Flora North America Program. Brittonia 25:119–134.

Morse, L. E. 1971. Specimen identification and key construction with time-sharing computers. Taxon 20:269–282.

———. 1974. Computer programs for specimen identification, key construction and description printing using taxonomic data matrices. Publ. Museum Mich. State Univ., biol. ser., 5(1):1–128.

Pankhurst, R. J. 1971. Botanical keys generated by computer. Watsonia 8:357–368.

———. 1974. Automated identification in systematics. Taxon 23:45–51.

Pettigrew, C. J. and L. Watson. 1973. On the identification of sterile Acacias and the feasibility of establishing an automatic key-generating system. Austral. J. Bot. 21:141–150.

Robinson, B. L. 1923. The need of monographic activity in American botanical taxonomy. Science 62:307–310.

Shetler, S. G., et al. 1973. An introduction to the botanical type specimen register. Smithson. Contrib. Bot. 12:1–186.

Soper, J. H. 1964. Mapping the distribution of plants by machine. Canad. J. Bot. 42:1087–1100.

———. 1969. The use of data processing methods in the herbarium. Anales Biol. Univ. Nac. México, Ser. Bot., 40:105–115.

Stuessy, T. F. 1975. The importance of revisionary studies in plant systematics. Sida 6:104–113.